A Framework for Modeling Human Spatial Information Processing in Command and Control Systems

Dr. Carsten Winkelholz, Mr. Michael Kleiber, Prof. Dr. Christopher Schlick
Fraunhofer-FKIE
Neuenahrer Str. 20
53343 Wachtberg
GERMANY

carsten.winkelholz@fkie.fraunhofer.de / michael.kleiber@fkie.fraunhofer.de / christopher.schlick@fkie.fraunhofer.de

ABSTRACT

One of the most important basic cognitive tasks of military operators the processing of spatial information. Soldiers involved in urban warfare need to be aware of the locations of red and blue forces. Local restrictions resulting from the topography of the urban battle space are taken into consideration when planning the next actions to complete a mission. Within a command and control system a number of tools exist to support the operator in processing spatial information. For the prospective design of human-machine interfaces of these tools in an engineering process, core principles of human performance need to be formulated into a computational theory. The proposed contribution will present a framework we developed to model human spatial information processing within a symbolic architecture of cognition. Within a sub-symbolic layer, spatial relations are represented as probability density functions modelling the uncertainty of their cognitive representation. Reasoning about spatial relations perceived from an external graphical display and their relation to internal cognitive spatial relations is reflected within the framework by the principle of Bayesian inference. In this way the cognitive processes of the human operator interacting with the visually perceived spatial information from an operational picture can be modelled and the graphical representation within the human-machine interface can be optimized. This contribution will discuss the theory for an application in an Anti-Air Warfare (AAW) scenario.

1.0 MODELING HUMAN SPATIAL COGNITION

Spatial perception and the geometry of visual space have been studied for more than a century. In the early days, scientists were interested in identifying a single geometry that maps Euclidean physical space into perceived visual space [1,2]. Later, the focus shifted to investigating the memorial space [3, 4, 5, 6]. Whereas detailed mathematical descriptions of the visual geometry have been developed for visual perception [2, 7, 8, 9], mostly normative models exist for cognitive maps. For instance, [6] proposed that locations are memorized in egocentric and allocentric coordinate systems. Allocentric coordinate systems define locations with respect to objects in the environment, which is consistent with the view that humans process spatial information in a mostly sequential fashion. The visual system scans and encodes a scene using a number of sequentially performed attention shifts, and the locations are recalled by sequentially retrieving spatial relations from memory that are related to the location in question. Therefore, it is reasonable to assume that spatial relations that shape the mental model are acquired during attention shifts. We believe that shifts of attention between several locations in space define the reference axes and planes of local allocentric coordinate systems within which the spatial relations are encoded. This assumption is consistent with the idea that locations are encoded by intrinsic frames of references [10, 11, 12, 13]. These intrinsic reference frames would result naturally from salient landmarks of the scene that attract attention. It should be possible to obtain a mathematical description of these spatial relations within the allocentric reference systems. In general, locations are described mathematically by scalar values. Since the neural system can only represent noisy values of scalar dimensions, it is reasonable for the cognitive system to

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **OCT 2010**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **A Framework for Modeling Human Spatial Information Processing in Command and Control Systems**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Fraunhofer-FKIE Neuenahrer Str. 20 53343 Wachtberg GERMANY**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

See also ADA564696. Human Modelling for Military Application (Applications militaires de la modelisation humaine). RTO-MP-HFM-202

14. ABSTRACT

One of the most important basic cognitive tasks of military operators the processing of spatial information. Soldiers involved in urban warfare need to be aware of the locations of red and blue forces. Local restrictions resulting from the topography of the urban battle space are taken into consideration when planning the next actions to complete a mission. Within a command and control system a number of tools exist to support the operator in processing spatial information. For the prospective design of human-machine interfaces of these tools in an engineering process, core principles of human performance need to be formulated into a computational theory. The proposed contribution will present a framework we developed to model human spatial information processing within a symbolic architecture of cognition. Within a sub-symbolic layer, spatial relations are represented as probability density functions modelling the uncertainty of their cognitive representation. Reasoning about spatial relations perceived from an external graphical display and their relation to internal cognitive spatial relations is reflected within the framework by the principle of Bayesian inference. In this way the cognitive processes of the human operator interacting with the visually perceived spatial information from an operational picture can be modelled and the graphical representation within the human-machine interface can be optimized. This contribution will discuss the theory for an application in an Anti-Air Warfare (AAW) scenario.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**14**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

memorize a location in the dimensions of a coordinate system by storing the values that are most suitable for an expected task. If locations in the vicinity of a landmark location need to be encoded with greater precision, the dimensions of a spherical coordinate system are a rational choice. This assumption is consistent with the fact that humans think about visual space in terms of distances and angles. In the following we describe a framework that is able to integrate the numerical mathematical description of spatial relations into a symbolic architecture of cognition. We used this framework already to model phenomena related to memorizing object locations in graphical structures [14] and for symmetry detection [15].

Symbolic Architectures of Cognition

Symbolic architectures of cognition describe cognitive processes by distinct steps. The current state of the environment and the cognitive system activates a rule that describes the actions and the state transitions of the cognitive system. ACT-R [16] is a popular architecture of cognition. Its essential feature is a sub-symbolic layer which influences the activation processes of rules and the state transitions of the cognitive system. The values of the parameters of the sub-symbolic process are variable and might adapt during the cognitive process, which enables the cognitive system to learn. The most elaborated module of ACT-R is the memory module. The memory is assumed to be a collection of symbolic entities $d^{(i)}$ called chunks. The probability of a successful retrieval and the retrieval time are determined by the activation of the memory chunks, which is calculated by a formula taking into account the time spans in which a chunk has already been retrieved, the strength of its association with the current goal and the similarity of the attributes in a request to the values in the attributes of a memory chunk. This central equation of ACT-R is given by:

$$a_i(t) = b_i(t) + \sum_j w_j s_{ji} + \sum_k u_k m_{ki} \qquad (1)$$

The base activation $b_i(t)$ decays logarithmically over time and increases each time the memory chunk is retrieved. The parameters s_{ji} reflect the frequency of how often chunk $d^{(i)}$ has been retrieved if the symbolic value of attribute ν_{jg} of the goal was identical to the current value. The parameter m_{ki} is the similarity parameter, and we think it could best be interpreted as the log-probability $m_{ki} = \ln(P(\nu_{kx} = \nu_{ki}))$ that the value in attribute ν_{kx} in the request x is identical to the value in the attribute ν_{ki} of the chunk $d^{(i)}$. Because in this case the ACT-R equation for the probability that a memory chunk will be retrieved factorizes to:

$$P_i \propto e^{b_i + \sum_k u_k \ln(P(\nu_{ki} = \nu_{kx}))} = e^{b_i} \prod_k e^{u_k} P(\nu_{ki} = \nu_{kx}) . \qquad (2)$$

The product in the factorized version expresses the probability that all attributes are equal to the attributes in the request. The parameters w_j and u_k are weighting factors reflecting the importance of a single attribute. To decide during a simulation which chunk will be retrieved from memory, noise is added to (1), and the random variable $A_i = a_i(t) + X + Y$ (2) is considered. The random variables X and Y are independent normal distributed with a mean of zero and a standard deviation σ_X; σ_Y. The value of the first random variable X is added when the chunk has been created. And the second one Y is added when $a_i(t)$ is re-evaluated. The memory chunk with the highest activation will be retrieved. If the activation of no memory chunk exceeds a threshold τ_a a failure will be retrieved. Because of the decaying of the base activation a memory chunk will be forgotten unless it is not frequently retrieved. The time needed for a successful recall also depends on the activation by the relation t ~ e^{-A}. The higher the activation, the faster a memory chunk can be recalled.

To model the variability of reproduced spatial relations from human memory such a sub-symbolic layer is also needed for a visual spatial module. Visual information is acquired by single shifts of attentions.

Reference Systems

The location of an object can only be identified within a frame of reference. In experimental psychology it is a well accepted procedure to divide the frames of references into two categories: An egocentric reference system, which specifies the location of an object with respect to the observer and an environmental (allocentric) reference system, which specifies the location of an object with respect to elements and features of the environment. According to Mou & McNamara [10] humans also use reference systems concerning the intrinsic axis of the object configuration. E.g. two salient objects on a plane create an axis that is used to specify the location of other objects. The most natural way to integrate this into the concept of attention is to consider the last two attended objects as an axis of reference. However, creating object-location memory chunks in this “semi-allocentric” reference system is less effort to the visual module because it only needs to keep track of two objects, whereas in the case of the pure allocentric reference system three objects are needed. Therefore in some situations the production system might be forced to use spatial memory chunks in the semi-allocentric system. In three dimensions three locations are needed to define a pure allocentric reference system. Three locations define a reference plane, whereby the cross-product of the two linear independent vectors that can be defined by the three locations determine a third axis for a right handed reference system. In general these three axes can be used to define the spatial relation of a fourth object to one of the three objects. Semi-allocentric reference systems for a three-dimensional spatial relation can be defined by replacing one of or both of the intrinsic reference axes with the viewing axis or the up-vector of the egocentric reference system. We think that humans represent, in their mental model, a spatial relation by its angle and the fraction of its length to the reference axes. This can be deduced by measuring the structure of the variability of reproduced object locations [16, 17, 18]. The introduction of object-relations based on three (four) objects is important for three reasons: First, it fits well with the concept of intrinsic axis in the object configuration as reported by Mou & McNamara [10]. Second the concept of angles is essential to most cognitive operations in geometric tasks. Third, it is the simplest percept for spatial memory chunks that allows reconstructing object locations, also if the whole configuration is rotated. The distance dimension is crucial, since there is no universal scale like for angles where any angle can be seen in relation to a complete rotation around 360°. The distance can be represented in respect to the length of the reference axis or in respect to the field of view. The representation in respect to the length of the reference axis is scale invariant and can later also be applied to a scaled environment (e.g., the screen is farther away). Representations of distances in respect to the field of view are considered in a semi-allocentric reference system. A semi-allocentric reference system is considered to use one location as the origin but to interpret direction and distance of a location relative to this point in an egocentric reference system. In our understanding, a distance is not perceived directly, but as a dimension of a location in one of these allocentric reference systems. For the model described in the following section it needs to be emphasized that if the distance is considered in the pure-allocentric reference system, the visual system needs to shift attention three times to assess a location—and therefore a relative inter-object distance. However, if the distance is represented in the semi-allocentric reference system only two attention shifts are needed (taking the attention shift to the first reference object into account).

The variances in recalled object-locations require the scalar values of a spatial relation to be noisy. To integrate noise into the location information of memory chunks the first question is how object-locations in different reference systems are represented in memory. [16] showed among other things, that the distribution of recalled locations supports the assumption that subjects imagine object-locations on a plane relative to a center in polar coordinates. We generalized this to use spherical coordinates in respect to model spatial cognition in three dimensions. This assumption has also some interesting implications on the representation of locations on a screen. Spherical coordinates are a system of curvilinear coordinates that are natural for describing positions on a sphere or spheroid. Generally ϕ is defined to be the azimuth angle around the polar axis, which is normally the up-vector, θ to be the zenith angle from the polar axis and r to be distance (radius) from a point to the origin. In the case of an allocentric reference system on a two dimensional screen this means, that if the three points p_{-2}, p_{-1}, p_0 were attended and p_0 has to be represented in a local allocentric reference system, the point p_{-1} defines the origin, the polar axis is given by (p_{-1}, p_{-2}), and the local spherical y-axis points orthogonal into the screen. For the semi-allocentric reference system on a screen, again p_{-1} is the origin, but the polar axis is parallel to the vertical axis of the screen and the x-axis is parallel to its horizontal axis. The next question is, if θ, ϕ, and r should be considered as single, independent memory chunks. Because it is impossible to imagine a distance without a direction and an angle without corresponding lines, it is reasonable to combine distance and angle as one percept in one memory chunk. Because of this argument, also in the case of the actual allocentric reference system the egocentric orientation of the reference system should be stored into the same memory chunk. This does not imply that the angle or the different dimensions of one chunk can not be separated later. In spatial reasoning often two angles have to be compared. But this can be handled as commands to the visual module of the cognitive system. Then, timing issues can also be considered, e.g. for the mental rotation of an actual allocentric reference system. In principle the spatial information of the semi-allocentric reference system is now also present in the chunk of an actual allocentric reference system. This might suggest discarding memory chunks of the semi-allocentric reference system. But as mentioned above, creating object-location memory chunks in this semi-allocentric reference system is less effort to the visual module and therefore in some situations needful.

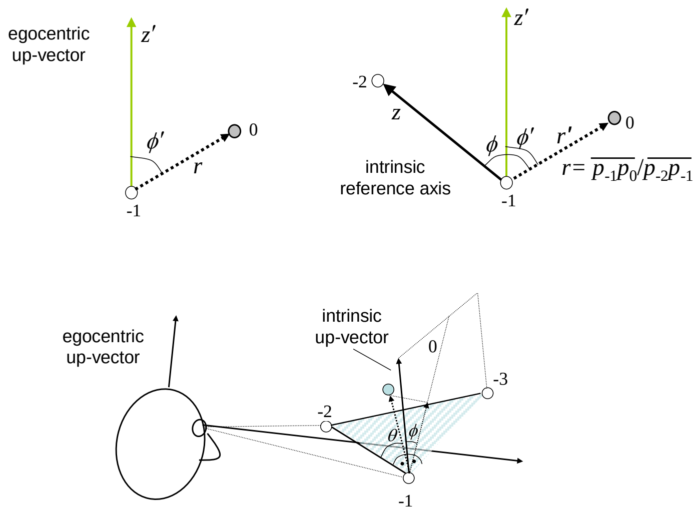

Figure 1: Different reference systems. The objects are attended in the order (p_{-3}, p_{-2}, p_{-1}, p_0).

Noise

Finally a spatial location is represented by $D(r,\theta,\phi, r',\theta',\phi',e_{rs})$, where r, θ, ϕ are the spherical coordinates as described above, e_{rs} indicates in which reference system r, θ, ϕ have to be interpreted, and r', ϕ', θ' are additional attributes for the actual allocentric reference system and additionally hold the polar and azimuth angle in the semi-allocentric reference system. The values of the spherical coordinates in the memory chunk are interpreted as random numbers distributed according to a truncated normal distribution $N(x, x_0, \sigma_x)$, with standard deviations (σ_r, σ_θ, σ_ϕ) corresponding to each dimension. The scalar value in the slot of the memory chunk indicates the mean x_0 of the distribution. Independent of whether the distance is measured compared to the field of view or to the length of the reference axis the ability of humans to discriminate them should obey the Weber-Fechner Law. Therefore, the noise σ_r should scale linearly with r_0. For later calculations it is easier to consider the quantity $f_r = r/r_0$ to specify the probability distribution. The noise model of f_r is again simply normally distributed with mean at 1 and a specific standard deviation σ_{fr}. In this way the noise σ_r is scale invariant. Therefore, if no scale transformations are to be considered the physical measurements of a distance can be used in the calculations. This is true regardless of whether the distance is represented as a fraction of the reference axis or as a fraction of the field of view.

Every time a location is to be encoded, it is decided if the perceived values for the location correspond to an already existing memory chunk. The posterior probability $P_{Di}=P(D_i=d^{(i)}|F_x)$ that the location of a feature F_x belongs to a memory chunk D_i and the probability P_0 that no appropriate memory chunk already exists, are given by

$$P_{D_i} = \frac{P(F_x \mid D_i)}{V^{-1} + \sum_i P(F_x \mid D_i)}, \quad P_0 = \frac{V^{-1}}{V^{-1} + \sum_i P(F_x \mid D_i)} \qquad (3)$$

The parameter V^{-1} describes the weight of a noisy background and it is assumed that single dimensions of the spatial relation are not correlated.

$$\begin{aligned} &P(F_x = (r_x, \theta_x, \phi_x, r'_x, \theta'_x, \phi'_x) \mid D = d(r, \theta, \phi, r', \theta', \phi')) = \\ &\qquad N(r_x, r, \sigma_r) N(\theta_x, \theta, \sigma_\theta) N(\phi_x, \phi, \sigma_\phi) N(r'_x, r', \sigma_{r'}) N(\theta'_x, \theta', \sigma_{\theta'}) N(\phi_x, \phi, \sigma_{\phi'}) \end{aligned} \qquad (4)$$

On the other hand, if an object-location is requested based on a memory chunk $d(r,\theta,\phi, r',\theta',\phi',e_{rs})$, the values are set to random values x according to (5). After the noise has been added to the location request, it is decided if the values are latched on possible features in the display. Therefore, the object-locations of all features $F_i(r_i, \theta_i,\phi_i,r_i',\theta_i',\phi_i')$ in question are calculated in the current local reference system corresponding to the reference system in the request. The probability P_{Fi}, that the location request is caught by feature F_i and the probability P_0 that it is not, are given similarly to (3) by

$$P_{F_i} = \frac{P(x \mid F_i)}{V^{-1} + \sum_i P(x \mid F_i)}, \quad P_0 = \frac{V^{-1}}{V^{-1} + \sum_i P(x \mid F_i)}. \qquad (5)$$

These equations express the posterior probability $P_{Fi}=P(F_i|x)$ that if a noisy location x from memory is given it results from the feature F_i. The likelihood probability functions $P(x|F_i)$ are the probability density according to if the feature F_i would have been the stimulus and are similar to (4). This process of encoding and reconstructing a location into a random number in memory is illustrated in Figure 2.

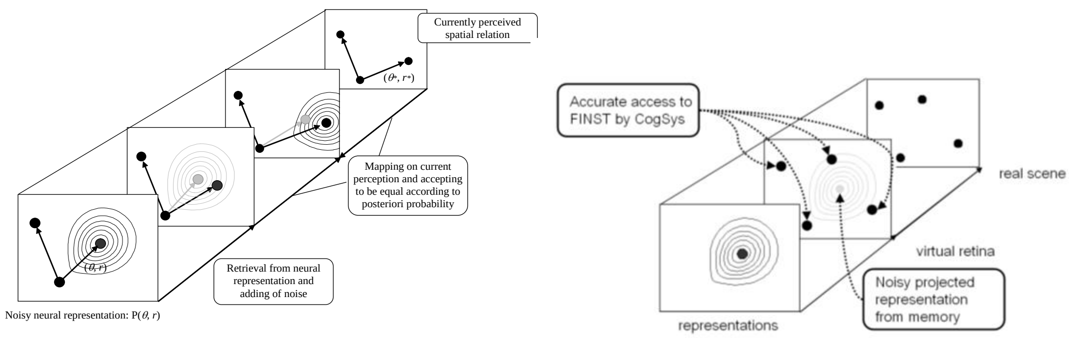

Figure 2: Perception, representation and reconstruction of a location.

The presented noise model has two interesting properties. First, because the truncated logistic distribution is asymmetric, the expected report of an object-location is biased away from the reference axis. This is the same effect as has been reported about categorical boundaries [16]. Second, for object-locations on a flat screen the values of ϕ, ϕ' are discrete $\phi,\phi'=(\pi/2,0,-\pi/2)$ and encode whether the object-location in question is on the left side, on the right side, or aligned, when facing into the direction of the reference axis. This is consistent with the assumption of interpreting the reference axis as a categorical boundary, where ϕ encodes the category.

From the results of our experiments we calculated for the noise parameter $\sigma_{fr} = 0.11$, which means that the length of a spatial relation can be reproduced by a human with a standard deviation of 11%. For the angle we got $\sigma_{\phi} = 6.3°$, which is 1.75% of 360° and $\sigma_{\theta} = 11.3°$ which is 3.1% of 360°. In general we try to use the same values for both allocentric reference systems. These values were measured for one specific view in 3D, where the viewing axis was along the reference axis. The experiments suggested that the values in 3D strongly depend on the view. In compromise with other data we suggest to use in the 2D case $\sigma_{fr} = 0.1$ and $\sigma_{\{\theta,\phi\}} = 8°$. However, all values show that directions can be memorized more efficiently than length. This is an aspect which could be an important design criterion of graphical displays. One example will be discussed below.

Visual Indexing

It is evident that subjects browsing a graphical layout structure encode environmental characteristics of object-locations, e.g. if an object is located on the border of a matrix. To encode such environmental features the cognitive system needs to attend objects nearby. The crucial point is that after some objects in the environment have been attended, attention needs to return to the object in question. If this return would depend on noisy spatial memory chunks, the strategy to encode environmental features might be highly counterproductive. At this point the concept of visual indexing, or FINST - FINger INSTantiation, [20] is needed. According to this theory the cognitive system has “access to places in the visual field at which some visual features are located, without assuming an explicit encoding of the location within some coordinate system, nor an encoding of the feature type”. Experiments suggest that the number of FINSTs in the visual system is limited to 4 to 5. In the visual module of ACT-R the concept of FINST is used to decide if an object has already been attended. Whenever an object is attended, a FINST is created. Because the number of simultaneously existing FINST is limited, any time a new visual object is attended the oldest FINST is removed to create a new FINST for the currently attended object. To implement environmental scan patterns, FINST needs to provide, additionally to the information that an object has already been attended, also information for accessing its location without, or at least with only minimal noise. In the visual module interface described below this has been accomplished by determining a visual index through the sequential position in the chain of attended locations. This index can be used in visual module commands to return (or avoid to return) attention to a particular location in the chain of attended locations.

2.0 MODELING EXAMPLES

In the following we will give some examples on how human spatial information processing can be modelled within the proposed framework.

Modeling State Assessment from Graphical Displays

The Fraunhofer FKIE developed a Tactical Situation Display (TDS) for Anti Air Warefare (AAW) scenarios with some innovative display elements (Witt et al. 2009). One of these new unconventional display elements is the polar display for fast survey over the track attributes. The polar displays were used to constitute in integrated form the track attributes that are crucial for threat classification, such as distance to the ownship (DST), altitude (ALT), speed (SPD), course (CRS), IFF-information (IFF) and ESM-emissions (ESM). For a single track attribute the polar display generates display proximity between the current attribute value and – taking into account predefined identification criteria – an un-critical attribute value on the attribute’s parameter beam.

A symmetric figure represents the a priori defined scenario knowledge. For instance, it is known in advance which friendly, neutral (civil) and hostile radar emissions (ESM emissions) are to be expected. Similar friendly and neutral IFF (identification friend foe) codes are defined a priori. If potentially threatening attribute values are detected the respective value indicator breaks the symmetry within the figure. The normal range of kinematic attributes like speed or altitude is defined in advance as a tolerance area. By connecting the current indicator values of single attributes in a polar diagram a figure is generated which integrates the single pieces of information on a higher level of abstraction. This figure by means of symmetry or a-symmetry forms a so-called emergent feature which helps to transfer the interpretation of information content to the perception phase of human information processing, i.e. direct perception: A symmetric figure indicates an un-critical airborne contact. In contrast, an easy to perceive a-symmetry emphasizes the criticality of the contact. The most relevant advantage of polar displays, in spite of the graphical aggregation of individual attributes, is that they assure that the single pieces of information are better noticeable and perceivable. Thus, in contrast to classical alarm displays polar displays are alarming and diagnostic at the same time, because the possibly symptomatic characteristic of a single parameter value can be noticed easily. Furthermore, under different parameter constellations the figure-forming aggregation of single attributes allows for the direct derivation of higher-level task-related manifestations. In contrast, the notification about several pieces of information on separated displays as mentioned above requires multiple mental transformations and comparisons.

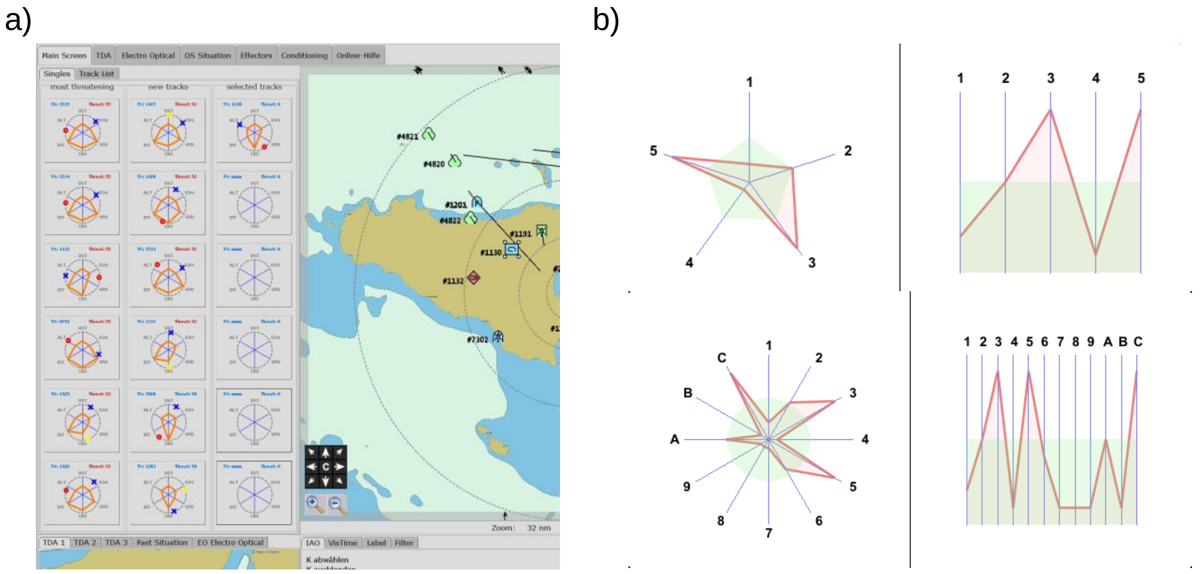

Figure 3: (a) The usage of polar display in a tactical situation picture of an AAW interface; (b) Comparison of the polar display to a linear display.

However, also alternative designs of displays exist that integrate the values of attributes into a figure. Figure 3 on the right shows the comparison of the same attributes displayed in two different designs for a graphical display. Both designs constitute figures that can directly be perceived and interpreted by the user. However, considering equally distributed values for the attributes, the characteristics of the figures that are constituted differ in both designs. The question is which design produces the more salient features that can be recognized more easily from figures stored in memory. This question can be discussed within the proposed framework, by examining the emergent angles and fractions of distances within the figures. These angles and fractions of distances are restricted by the design of the display (see Figure 4).

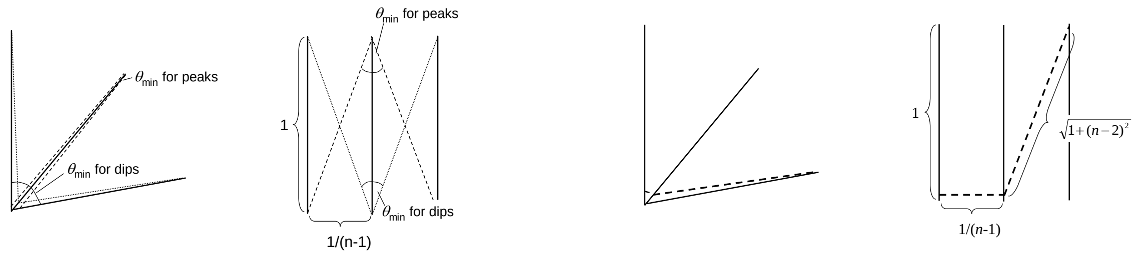

Figure 4: Illustration of the restrictions the design of the displays put on values of attributes of spatial relations.

In case of the linear display the angle ranges $[2 \cdot \tan^{-1}(1/(n\text{-}1)),180°]$ within peaks and dips, where n is the number of axes in the diagram. In a polar display the ranges differ between peaks and dips. The angle within a peak can theoretically become zero as a minimum and 180°-360°/n as a maximum. For peaks pointing inward (dips) the smallest angle that can emerge is $2 \cdot 360°/n$. Therefore the ranges of angles for peaks and dips are $[0,180°\text{-}360°/n]$ and $[720°/n, 180°]$. For the fraction of distances of one arm to another in a linear display the range is $[(n\text{-}1)\sqrt{(1+(n\text{-}1)^{-2})},\ 1/((n\text{-}1)\sqrt{(1+(n\text{-}1)^{-2})})]$. In case of the polar display theoretically the fraction of length can range from 0 to infinity. However, we assume here that it is bounded by the r = 10, and we think that also σ_{fr} = 0.1 only holds for r smaller than 10 or larger than 1/10. Figure 5 visualizes the ranges of angles and fractions of lengths. As can be seen from Figure 5 the linear display provides a larger range of angles, especially if the number of axes in both displays is low. For the fraction of length the polar display provide a larger range of fractions that could emerge and could be discriminated by an operator.

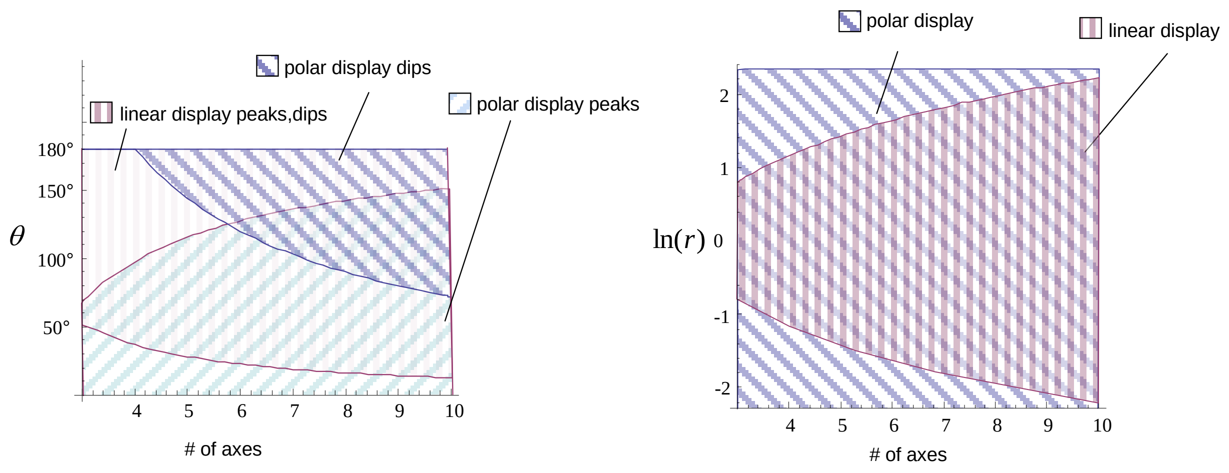

Figure 5: Comparison of the ranges for the values of the spatial attributes between polar and linear display.

In a common task an operator needs to classify an alert that is signalled by the constituted figure within a display. For this he needs to compare the figure perceived with the figures from memory. The operator will extract salient features from the figure and will try to recall a figure possessing this salient feature from memory. These salient features are spatial relations in nature. As discussed above humans are limited in recalling spatial relations and this process is noisy. To identify a figure he will compare different features to come to a decision, about the meaning of the displayed alert. If one salient feature is very important for the characteristics of one figure he might have stored this by more than one single spatial relation. We will not discuss a complete model here. The differences in performance between both display designs result from the different ranges of spatial relations that can occur in the figures displayed. As shown above, these ranges depend on the number of axes, and the linear display seems to be advantageous for less number of axes. However, the noise from memory is different for different attributes of the spatial relations and this needs to be considered within the assessment of both display types. To get a first model based assessment for both display types we consider the step in recalling the first salient feature of a displayed figure. For this two spatial relations are needed: one for the shape of the salient feature and one for its position within in the display. Both spatial relations are shown for both display types in Figure 6.

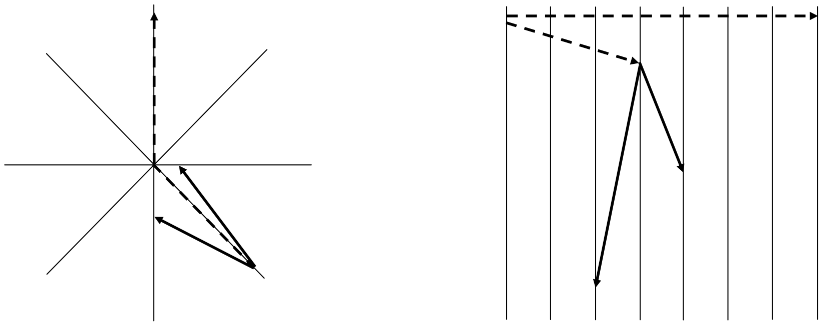

Figure 6: Encoded spatial relation for the relative location for one salient feature in the two different displays.

As a measure of performance we calculate the mean probability that the shape of the salient feature can be recalled from memory successfully and if the associated position which is recalled from memory will correctly be recognized. Actually the performance in both steps depends on how well the spatial relations can be discriminated from other spatial relations that are associated with figures in the display. The overall performance will be a weighted sum of the performance in both steps. Both steps could also be merged into one single recall for one memory chunk holding both relations, for the shape and its location. In this case, again the overall performance would be a weighted sum of the ability to discriminate the spatial relations from others. In fact we are not sure on how the weights have to be chosen or if they could be concluded from the ACT-R theory. However, the separation into these two steps clarifies the different contributions of the shape of the feature and its location.

For the first step we consider that only this salient feature is in memory and it only needs to be distinguished from the background noise. The expected probability $P(D_{Fx} = d^{(Fx)}|F_x)$ that the framework will propose the salient feature F_x perceived to be equal to the same noisy salient feature $d^{(Fx)}$ in memory can be calculated by integrating eq. (3) over all possible attributes of the spatial relations that constitute the feature within the display.

$$\left\langle P_{D_{Fx}} \right\rangle = \int_{\Omega(Fx)} \frac{P(F_x \mid D_{Fx})}{V^{-1} + P(F_x \mid D_{Fx})} P(F_x)\, dF_x , \qquad (6)$$

To simplify calculations we consider that the displayed attributes on the axes are distributed in a way, that the probability density functions is uniform $P(F_x)=1/\Omega(F_x)$, which can also be interpreted as the background noise $V^{-1} = P(F_x)$. Furthermore, it is assumed that exactly the same spatial relation has to be retrieved from memory that has been stored previously. Now the conditional probability density functions within the integral only depends on the standard deviation as a parameter.

$$P(F_x \mid D_{F_x}) = N(x_0, x_0, \sigma_x) = g(\sigma_x) \qquad (7)$$

The variable ϕ, which is in the 2D case discrete, determines if the feature is a peak or a dip. We assume that there is no noise in the cognitive system about the categorization of a feature into a dip or a peak. Therefore, the consideration of ϕ causes (7) to be separated into a weighted sum of two integrals: one for the peaks and one for the dips. For the linear display the ranges of spatial attributes are the same for peaks and dips, as is also the probability of their occurrence. For the polar display it must be considered, that there are only dips for $n > 4$ and that the ranges of the attributes differ between dips and peaks. In the polar display the probability of the occurrence of peaks and dips differ and correlate with the ranges of the angles. In the same way should the probability mass of the fractions of distances be equally distributed on the interval [1/*n*, 1] and [1, *n*]. This is achieved by taking the logarithm of *r*, which has also already been done in Figure 5 for illustration purposes. Putting it all together results in Figure 7. It shows the mean probability that the shape of one salient feature can be discriminated from the background noise and can successfully be recalled as a function of the number of axes. As can be seen an increasing number of axes improves the recall probability of the shape of the salient feature. The linear display shows, in this aspect, a better performance than the polar display which reflects the larger range of angles. However, it should be stressed, that in a real ACT-R simulation this probability also depends on the base activation level for the memory chunk. Furthermore, we omitted by this calculation, that in a polar display the direction of the peak might inherently be encoded within one memory chunk, because the allocentric reference system have different orientations depending on which axis the peak is. This is not the case for the linear display, where the orientations of the allocentric reference systems are always the same and aligned with the egocentric up-vector.

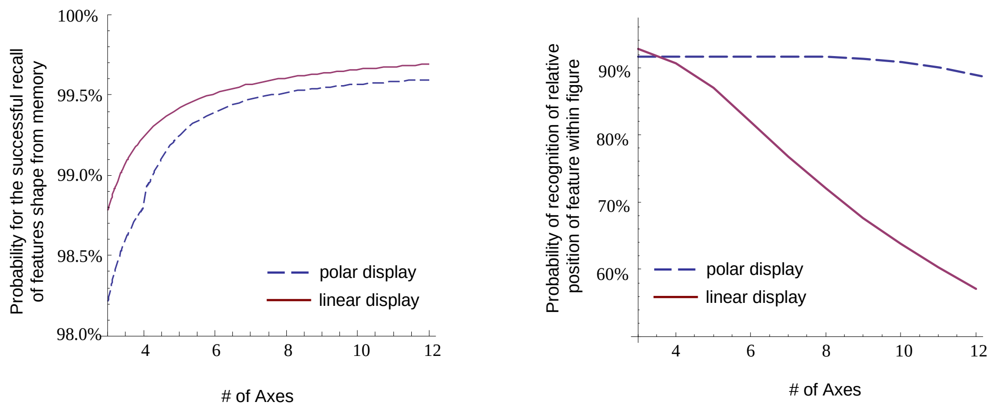

Figure 7: Predicted performances for the recall of the feature's shape and for the recognition of its relative position within the figure.

After the memory chunk of one salient feature has been retrieved successfully it is plausible to assume that the cognitive system has stored an associated memory chunk for the position of the salient feature within the figure. The cognitive system will than validate this position of the salient feature with the perceived position. The probability that the correct position of the salient feature will be accepted can be calculated within the framework by:

$$P(F_{x_0} \mid D_0 = d^{(0)}) = \int_{\Omega(x)} P(F_{x_0} \mid x) P(x \mid D_0 = d^{(0)}) dx = \int_{\Omega(x)} \frac{P(x \mid F_{x_0})}{V^{-1} + \sum_{i=1}^{n} P(x \mid F_{x_i})} P(x \mid D_0 = d^{(0)}) dx$$

The integral 'simulates' the noise that is added to the spatial attributes stored in the memory chunk. The possible features on which the location from the memory chunk could be latched are the discrete locations on the axes and therefore are given by: $P(x_0 \mid F_{x_i}) = N(x_0, x_i, \sigma_x)$. Again it is assumed that the correct feature spatial relation has been recalled from memory: $P(x \mid D_0 = d^{(0)}) = N(x, x_0, \sigma_x)$. The resulting predicted performance is shown in Figure 7 on the right side. For the recognition of the position of the feature within the display the polar display clearly outperforms the linear display. This can be explained by the fact that the standard deviation of the noise for angles is smaller than for distances. The linear display strongly decreases with a higher number of axes, while in the polar display it seems to stay constant up to twelve axes.

Modeling the Assessment of the Operational Picture from Different Views

To make tactical decision the operator needs an operational picture. In general he needs to know where different entities are located in relation to the ranges of effectors and sensors and how these locations will change in time. The operator needs answers to questions like: which entity will be next within reach of the effectors; which sensors and effectors will I be able to use; where is the best location to hit the enemy aircraft without endanger civilian entities in the proximity. All these questions can be answered by algorithms implemented in the combat computing system with high accuracy. However, as long as the operator needs to make these queries sequentially or has to read the answers of the system in form of lists, performance in making the right decisions will be hampered. An external visualisation of the spatial layout of the operational picture should enable the operator to get answers to some of his queries that refer to spatial relations and time evolution intuitively by perception. To handle all spatial relations numerically in a list is impossible, because the number of possible spatial relations increase strongly (~ n^2) with the number of entities in the operational picture. However, also the derivation of spatial relations from a spatial layout can be biased and can only be derived by humans up to certain level of accuracy.

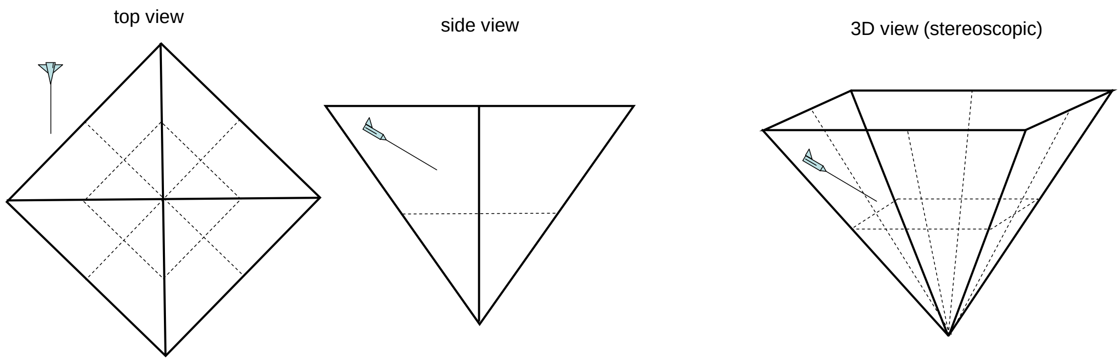

Figure 8: Scenario: Decide if the aircraft will enter the range volume of a specific sensor if continued its course.

The physical parameters of sensors, effectors and tracks are three dimensional in nature. The choice of the views on the operational picture strongly influences performance of different cognitive tasks the operator needs to perform to extract the information he needs for his decision-making process. By modelling the process of information acquisition from the display, not only different visualization approaches can be compared prospectively, but the model can also give clues about how a single interface design can be improved. One short example on how a certain view can be assessed by the proposed framework is illustrated in Figure 8 and will be discussed in the following.

One basic question an operator is interested in to query from the operational picture is, if an aircraft will come into the range of one specific effector. This information can only be retrieved if the spatial relations are imagined in three dimensions. To simplify argumentation we assume in the example of Figure 8 that the shape of the range-volume is a pyramid standing on its top. To infer if the aircraft will enter or pass the volume, the edge in proximity to the projected flying path is critical. One solution to decide if the flight path crosses the volume is to derive if the flight path runs in front or behind the edge, when the scene is looked at from a viewpoint where the edge in question points towards the observer. To do so, the speed vector of the aircraft is projected in its length beyond the edge. Then, an arbitrary location on the edge is attended and the attention is directed back to the location of the aircraft afterwards. In this way an allocentric three dimensional reference system has been created by two axes. As explained in the theory section a third “up”-axis is inherently defined by both these axis. Now the cognitive system simply needs to decide if this inherently created “up”-axis of the local reference system has an up or a down component. More generally the operator could also compare the “up”-axes of this local reference system with the “up”-axes of the reference system that can be constituted by the edges at the base of the pyramid. This is illustrated in Figure 9.

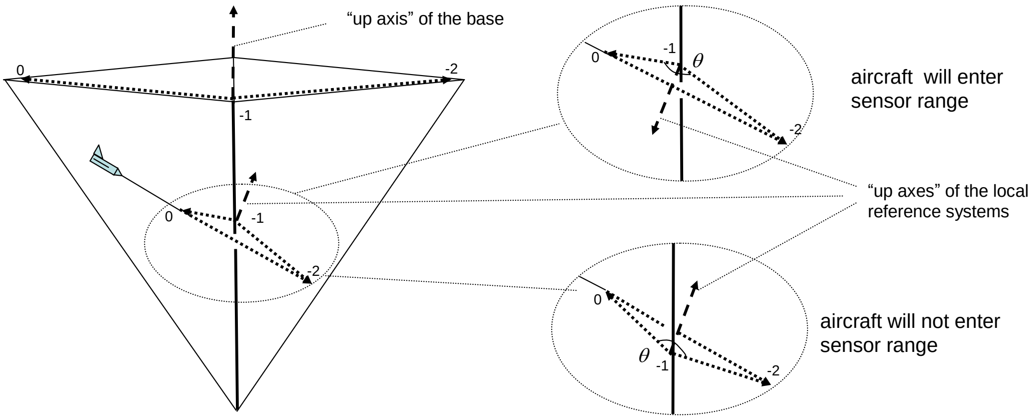

Figure 9: Cognitive steps to decide if the aircraft will enter the volume of the radar range.

The imaginary picture might be constituted from a stereoscopic 3D-visualization or multiple 2D views. In a stereoscopic 3D view the operator only needs to project the flight path by extending the speed vector in his imaginary picture. All other locations needed to constitute the reference system can be retrieved by perception. If the operator uses multiple 2D views he needs to memorize one spatial relation and to transform it into a second view. There are two drawbacks if a spatial relation needs to be reconstructed from memory. Beside that its reconstruction is noisy it possibly needs to be retrieved multiple times, if the number of FINSTs are not sufficient to keep one of it constantly on the constructed location. Whenever this FINST is needed, the spatial relation needs to be retrieved from memory again, which is time-consuming. In the example above the operator could memorize from the top view the relative location of the aircraft to the center axis of the volume. To construct this location of the aircraft in the side view, he than attends in the side view one location on the center axes in the height of the aircraft and add the rotated spatial relation from memory to this location. This imaginary location he needs to memorize, since he needs to perform another rotational task and he will not be able to keep attention constantly on this imaginary location. The other rotation task is needed to reconstruct the imaginary edge in the three dimensional space of the side view. Only if both imaginary spatial relations have been created by transforming them from memory, they can be tested as described above. As can be seen by this brief model description one query to multiple 2D views needs a chain of multiple noisy steps to perform the task. It is even arguable, if the number of FINSTs is sufficient to complete the task. However, a detailed analysis by a detailed model is ongoing work. But this short discourse shows how the framework might help to asses the limits of an operator to perform such geometric tasks. In a stereoscopic 3D view the task can be completed, because for some of the locations that are needed to constitute the reference systems, attention is caught by the real stimulus and variability of the attended locations are lower. But some steps are still noisy and this produces inaccurate responses. The probability of inaccurate responses in dependence on the operational picture and the viewpoint could be predicted by the model.

3.0 CONCLUSIONS & OUTLOOK

This article proposes a framework to model human performance for spatial reasoning tasks. However, even for simple tasks the modeller needs to make some assumptions about how operators guide their attention to encode the visual space into memory. However, if it is true that operators are equally restricted as it is described within the framework, special expertise in spatial reasoning tasks can be trained by learning to chose the appropriate reference system for one specific environment. The framework could be used in future work to analyse how experts solve specific problems in spatial reasoning and transfer the knowledge to novices.

Overall we feel that more basic research is needed to confirm the parameters for the noisy representation of basic spatial relations in memory. The measurement of these basic variability parameters is difficult, because it is difficult to control which reference locations participants will use to encode one spatial location within a laboratory environment. For example any dot represented within a display could be encoded by using any of the corners of the screen. Also an unsolved question is, if or how the noise of basic spatial relations increases in time. Actually, performance in recalling locations or an operational picture decreases in time. But this does not necessarily mean, that noise of one spatial relation in a single memory chunk increases. It could also be explained by a model in which the accuracy of a single spatial relation is relatively stable in time, but some of multiple memory chunks of spatial relations that were used to encode the location of one single object could not be retrieved anymore. Because some spatial relations to refine the spatial location of one object are missing, accuracy in reconstructing the object location decreases

Finally, we think that the proposed framework will be useful to understand limitations of humans in processing spatial information. In the first stage this will bring attention to possible problems in visualizing such information. In a second stage it could also be used for a model based assessment of different HMI designs.

REFERENCES

[1] Helmholtz, H.L.F. (1867). Handbuch der physiologischen Optik (1st ed.). Leipzig: Voss.

[2] Luneburg, R.K. (1947). Mathematical analysis of binocular vision. New York: Oxford University Press.

[3] Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. Cognitive Psychology, 10, 422–437.

[4] Byrne, R.W. (1979). Memory for urban geography. Quarterly Journal of Experimental Psychology, 31, 147–154.

[5] Tversky, B. (1981). Distortions in memory for maps. Cognitive Psychology, 13, 407–433.

[6] McNamara, T.P. (2003). How are the locations of objects in the environment represented in memory? In Freksa, C., Brauer, W., Habel, C., & Wender, K. (Eds.), Spatial cognition III: Routes and navigation, human memory and learning, spatial representation and spatial reasoning. (pp. 174–191). Berlin: Springer.

[7] Blank, A.A. (1957). The geometry of vision. British Journal of Physiological Optics, 14, 1–30.

[8] Indow, T. (1974). On geometry of frameless binocular space. Psychologia, 17, 50–63.

[9] Heller, J. (1997). On the psychophysics of binocular space perception. Journal of Mathematical Psychology, 41, 29–43.

[10] Mou, W., & McNamara, T.P. (2002). Intrinsic frames of reference in spatial memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28, 162–170.

[11] Schmidt, T., Werner, S., & Diedrichsen, J. (2003). Spatial distortions induced by multiple visual landmarks: How local distortions combine to produce complex distortion patterns. Attention, Perception & Psychophysics, 65, 861–873.

[12] Schmidt, T. (2004). Spatial distortions in visual short-term memory: Interplay of intrinsic and extrinsic reference systems. Spatial Cognition & Computation, 4(4), 313–336.

[13] Schmidt, T., & Lee, E.Y. (2006). Spatial memory organized by environmental geometry. Spatial Cognition & Computation, 6(4), 347–369.

[14] Winkelholz, C., & Schlick, C. (2007a). Modeling human spatial memory within a symbolic architecture of cognition. In Barkowsky, Th., Knauff, M., Ligozat, G., & Montello, D.R. (Eds.), Spatial cognition V: Reasoning, action, interaction. (pp. 229–248), Berlin: Springer.

[15] Winkelholz, C., & Schlick, C. (2007b). Bridging psychophysics and cognitive engineering in visual perception, Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 2520–2527.

[16] Huttenlocher, J., Hedges, L.V., & Duncan S. (1991). Categories and particulars: Prototype effects in estimating spatial location. Psychological Review, 98(3), 352–376.

[17] Winkelholz, C., Kleiber, M., Schlick, C.M. (2010a). Analysis of the Variability of Three-Dimensional Spatial Relations in Visual Short-Term Memory. Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society 2010. Mahweh, NJ: Lawrence Erlbaum Associates

[18] Winkelholz, C., Kleiber, M., Schlick, C.M. (2010b): Modeling human performance in processing basic three-dimensional spatial relations. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC 2010).

[19] Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. Psychological Review 111(4), 1036–1060.

[20] Pylyshyn, Z. W.: The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. Cognition 32 (1989) 65-97

[21] Witt, O.; Grandt, M.; Küttelwesch, H. (2009): Direct Perception Displays for Military Radar-based Air Surveillance. In: Engin. Psychol. And Cog. Ergonomics, HCII 2009, LNAI 5639 proceedings (S.606-615). Springer.